U.S. DEPARTMENT OF COMMERCE
NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION
NATIONAL WEATHER SERVICE
SYSTEMS DEVELOPMENT OFFICE
TECHNIQUES DEVELOPMENT LABORATORY

TDL Office Note 77-4

TESTING AND IMPLEMENTATION OF MOS MAX/MIN FORECAST EQUATIONS DERIVED FROM EXTENDED RANGE PE FIELDS

J. Paul Dallavalle and Gordon A. Hammons

March 1977

Testing and Implementation of MOS Max/Min Forecast Equations Derived From Extended Range PE Fields

by

J. Paul Dallavalle and Gordon A. Hammons

1. INTRODUCTION

The National Weather Service (NWS) currently produces automated forecasts of maximum and minimum temperature (max/min) at 228 stations in the conterminous United States. This objective guidance is formulated from linear regression equations that were developed by the Model Output Statistics (MOS) technique (Glahn and Lowry, 1972). In brief, the method correlates climatic terms, observed data, or model forecast fields interpolated to a station with an observed predictand at that same station. For max/min temperature forecasting, ten predictors are chosen in a forward, stepwise screening method. The predictand is the observed calendar day maximum or minimum, depending on the particular projection. During the 0000 GMT cycle, we make MOS forecasts that are valid for projections approximately every 12 hr from 24 to 60 hr after 0000 GMT. Thus, the first projection forecast is for the first day's max, the second is for the second day's min, and so forth. Analogously, at 1200 GMT, forecasts of the max/min temperature are made out to approximately 60 hr with the first projection prognosis being for a min, the second for a max, and so on. All of these forecasts are made from observations, climatic terms, or model fields which are valid up to 48 hr after the initial model time.

During the 0000 GMT cycle, the NWS has also been making a 72-hr objective forecast of the max prepared by the older perfect prog equations (Klein and Lewis, 1970). These forecasts required 60-hr output from the PE model and were available for 126 stations.

Previously, Hammons et. al. (1976) found that the 60-hr MOS forecasts of the min from 0000 GMT were inferior to perfect prog forecasts run as a control. Verifications done at the 126 stations common to both systems indicated that this problem was particularly bad during the 6-month cool season (October-March). However, when we restratified our developmental data to 3-month seasons (op. cit.) we substantially improved the 60-hr forecast of the min. Yet, we were not using any PE fields valid beyond 48 hr after 0000 GMT.

Since October 1972, the Techniques Development Laboratory (TDL) has been archiving various PE model forecast fields that are valid 60, 72, or 84 hr after 0000 GMT. Having accumulated 4 years of data, we felt that we could derive new 60- and 72-hr forecast equations for the 0000 GMT cycle. These new equations would be based on 48-, 60-, 72-, and 84-hr predictors stratified by 3-month seasons. As an added benefit, this new development would increase the station coverage at 72 hr by over 100 stations.

This note describes our experiments with the 60- and 72-hr MOS forecasts. After evaluating our results, we implemented new 72-hr MOS forecast equations for the winter (December-February) season on December 14, 1976. However, there was no improvement with the new 60-hr equations, so those were not implemented.

2. EXPERIMENTAL PROCEDURE

There were 4 years of PE model data available for this experiment. We decided to test on the winter season (December-February) since differences in the various techniques should be most evident in this period. For experimental purposes, 3 seasons (1972-75) were used as developmental data (approximately 230 cases); the 1975-76 winter was reserved as an independent sample. Both 60- and 72-hr equations were derived and tested for a selected set of 49 stations (Figure 1). At 72 hr we also compared the MOS forecasts to the operational perfect prog forecasts for the same stations and days.

Table 1 indicates the predictors that were offered at the 60- and 72-hr projections. The older, operational 60-hr predictor list is included for comparison. There are several distinct differences between the two 60-hr predictor lists. In the older list, a number of 24-hr model fields were used as potential predictors. Additionally, many of the forecast fields were smoothed by a 25-point space filter. Finally, no PE fields beyond 48 hr were considered. In contrast, the experimental 60-hr list contains no 24-hr predictors and no 25-point space smoothings, but uses predictors valid at 48, 60, and 72 hr after 0000 GMT. For the 72-hr MOS forecasts, potential predictors verify 60, 72, or 84 hr after 0000 GMT. Fields are often smoothed by a 5-point or 9-point space average.

For both the 60- and 72-hr projections, we offered several new predictors. Among these fields were the boundary layer vertical velocity, the boundary layer relative humidity, the relative humidity from the top of the boundary layer to 720 mb, the relative humidity from 720 to 490 mb, the temperature advection at 850 mb, and the 500-mb geostrophic vorticity advection.

Using the operational predictor list given in Table 1, we derived 60-hr min equations for 49 stations from 3 years (1972-75) of data. Forecasts based on these equations will be summarized under column "B" in the following results. We also made forecasts from last winter's operational equations; these equations had been developed previously from 6 years of data (1969-75) and the operational predictor list. Corresponding results will be given under column "A". Finally, we used the new predictor list (Table 1 - Test) to derive 60-hr min test equations from the 3-year data sample. The forecasts resulting from these equations will be labeled "C". By making three sets of 60-hr forecasts it is possible to determine both the effects of the new set of predictors (by comparing B and C), and the effects of using different dependent samples (by comparing A and B). The next section will show the advantages of this procedure.

3. RESULTS

Table 2 lists the standard error of estimate and the reduction of variance averaged at 49 stations for the three sets of 60-hr equations. These figures are based on dependent data and reflect how well the linear regression equations fit the developmental sample. For equation sets B and C there were 3 years of dependent data (1972-75), while for set A there were 6 years of data (1969-75). A similar set of statistics is given in Table 2 for the 72-hr equations. In this case, however, we developed equations on 3 years of dependent data (1972-75), tested them on independent data, and then redeveloped the equations from a 4 year sample (1972-76). In Table 2 we have included the average standard error at 49 stations from the 72-hr forecast equations on both developmental samples.

The average standard error on the dependent data (Table 2) for the 3-year 60-hr equations (set B) was substantially less than the standard error for the operational equations (A) on six years of developmental data. Adding new predictors and extended projections and developing equations on 3 years of data (C) resulted in a smaller standard error (by 0.5°F) than found with the operational equations (A). However, the standard error with C was only 0.1°F less than the error in set B. The implication is that forecast equation sets B and C fit the dependent data better (smaller standard errors) than set A primarily because the developmental sample was smaller (3 years as compared to 6). The new predictors (set C) contributed little additional information to set B. For the 72-hr MOS equations, the average standard error at 49 stations for the 3-year sample was 6.4°F. This increased to 6.7°F when another year of data was added and the equations were rederived from a 4-year sample.

We then used the 60-hr (all three sets) and 72-hr equations (developed on a 3-year sample) to make forecasts on independent data (December-February, 1975-76) at 49 stations. Verifications (Table 3) demonstrated a dramatic change in the relative merits of the three sets of 60-hr equations. On independent data, the forecasts from the operational equations (A) had smaller mean absolute errors than either of the equation sets (B, C) developed from 3 years of data. In fact, set A was 0.3°F more accurate in mean absolute error than the equations (C) that used extended-range predictors. Set C, however, was still 0.1°F mean absolute error better than set B. Thus, the extended predictors helped somewhat, but not enough to overcome the benefits of the larger dependent sample.

The biases (Table 4) also support the superiority of the currently operational 60-hr forecasts. Although sets A and C both have a mean algebraic error of 0.0°F, the operational forecasts (A) had fewer errors in the extreme categories than did the experimental forecasts (C).

At 72 hr, we compared the perfect prog forecasts to the temperatures predicted by the MOS 72-hr equations developed from 3 years of data. As before, the independent data came from the period of December-February, 1975-76. Although the differences in the mean absolute errors between the two systems varied greatly from one station to another, the perfect prog averaged slightly better (Table 3) in mean absolute error than the MOS forecasts. The mean algebraic error (Table 4) of the perfect prog forecasts was also much less (-0.3°F) than that of the MOS forecasts (-1.5°F). Similarly, the error distribution (Table 4) of the MOS forecasts was more skewed than that of perfect prog, although, overall, there were less large forecast errors (by 16) in the MOS system. We hypothesize that the larger developmental sample (18 years) and shorter stratification (2 months) make the perfect prog equations comparable in accuracy to the MOS forecasts at 72 hr. Also, the perfect prog equations are probably more stable than MOS.

4. OPERATIONAL IMPACT

Our results showed that at 60 hr we would likely not improve our present system by implementing new equations based on the limited sample of long-range predictors. However, for 72 hr we concluded that the benefits of increasing the station coverage, as well as obtaining a consistent, complete MOS system, compensated for the slight deterioration in the new forecasts made from extended predictors. Furthermore, the wide variability in results from station to station indicated that the small forecast degradation did not apply to all stations. Therefore, in the fall of 1976 we rederived the 72-hr MOS equations (winter season) for all 228 stations where data were available. We added the season used previously as an independent test to the dependent sample so that for any one station there were approximately 320 cases. The larger sample should make the resulting forecast equations more stable than those produced in our tests. Because of computational problems, we removed the boundary layer wind divergence and the 850-mb temperature advection from the list of potential predictors. Finally, we lacked sufficient data to develop equations for Zuni, New Mexico and Dallas, Texas. Thus, MOS forecast equations for the 72-hr projection were derived for only 226 stations. The new 72-hr MOS forecasts became operational in mid-December 1976 as part of the "final" FOUS22 teletype message (National Weather Service, 1976).

5. DEVELOPMENT OF THE WINTER SEASON OPERATIONAL EQUATIONS

For the rederived 72-hr winter season equations, the standard error averaged 6.6°F and ranged from 2.8° F at Key West, Fla. to 10.0°F at Dodge City, Kansas. The figures are consistent with extrapolations of the standard errors of the earlier winter projections (Hammons et. al., 1976). The reduction of variance averaged 64% and ranged from 42% at Stockton, Calif. to 76% at Billings, Montana.

The most important quantities for predicting the winter 72-hr max are listed in Table 5. As we have found in other work, the PE 850-mb forecast temperature is the most important predictor. In addition, the climatic (cosine) terms are significant because the model has less skill at the longer projections. The other important predictors include the temperature field in the lower atmosphere (1000 mb, boundary layer potential temperature, and the 850-1000 mb thickness), low level winds, and low-level relative humidities which, hopefully, indicate the presence of clouds.

6. DEVELOPMENT OF THE 72-HR SPRING SEASON OPERATIONAL EQUATIONS

We have now derived 72-hr max temperature equations for the 0000 GMT cycle in the spring (March-May) season. The dependent sample contained 4 years (1973-76) of data--approximately 325 cases. Unlike the winter derivations, we included the boundary layer wind divergence and the temperature advection at 850-mb as potential predictors in the spring development. Other than this change, the derivations in the two seasons were analogous. Equations were again derived at 226 MOS temperature stations. No equations are available for Zuni, New Mexico or Dallas, Texas.

In this particular derivation, we did no testing on independent data. The standard error on the dependent data averaged 5.9°F at 226 stations and ranged from 1.8°F at Key West, Florida to 8.1°F at Rapid City, South Dakota. This standard error is less than that in the winter season, as one would expect, since the variability of the predictand is less in the spring than in the winter. Furthermore, the 72-hr max standard error for the spring season follows the trends established in the early projections for the same season (Hammons et. al., 1976), namely, that the standard error for the max increases much more rapidly than that for the min in the spring. For the 72-hr spring equations, the reduction of variance averaged 75% and ranged from 44% at Santa Maria, California to 88% at International Falls, Minnesota.

Table 5 lists the important predictors for the spring 72-hr max equations. Again, the 850-mb temperature predicted by the primitive equation model is the quantity most likely to be used in forecasting the maximum temperature at the surface. Additionally, the climatic terms (cosine day of year, sine 2* day of year) are important predictors as they mirror the temperature curve during the spring transitional season. Other significant quantities include the atmospheric temperature structure (850-1000 mb thickness, 500-1000 mb thickness, boundary layer potential temperature), low-level winds (boundary layer U wind, 850-mb temperature advection), and some indication of the predicted cloudiness (layer 2 relative humidity, 500-mb geostrophic vorticity advection).

7. CONCLUSIONS

The results presented in this paper were used primarily to indicate the optimum operational configuration with our available data. However, we believe that these results also have important implications for our future work. We have shown the danger of using standard errors of estimate as the sole criterion for choosing a "best set" of forecast equations, especially with a small data sample. Panofsky (1958) estimated that in any meteorological sample of N consecutive days, there may be only N/7 to N/3 days of independent data. Consequently, we may, on occasion, be "overspecifying" our dependent sample by accounting for chance relationships between predictor and predictand (see Panofsky and Brier, 1958, pp 95 and 177). We, thereby, obtain standard errors that are too optimistic, that is, they do not give a realistic estimate of the relative accuracy one might expect on an independent sample.

When more data are available, we plan to rederive equations for the 60-hr min. The results presented previously indicate that we may make some improvement in these 60-hr forecasts if we have sufficient cases of the extended-range predictors. Further testing will be necessary.

In the future, we will also derive 72-hr forecast equations for the summer and fall seasons.

8. REFERENCES

Glahn, H. R., and D. A. Lowry, 1972: The use of model output statistics (MOS) in objective weather forecasting. J. Appl. Meteor., 11, 1203-1211.

Hammons, G. A., J. P. Dallavalle, and W. H. Klein, 1976: Automated guidance based on three-month seasons. Mon. Wea. Rev., 104, 1557-1564.

Klein, W. H. and F. Lewis, 1970: Computer forecasts of maximum and minimum temperatures. J. Appl. Meteor., 9, 350-359.

National Weather Service, 1975: Revision of FOUS12 and FOUS22. NWS Tech. Proc. Bull., No. 150, 8 pp.

Panofsky, H. A., and G. W. Brier, 1958: Some applications of statistics to meteorology. The Pennsylvania State University, 224 pp.

Reap, R. M., 1972: An operational three-dimensional trajectory model. J. Appl. Meteor., 11, 1193-1202.

Shuman, F. G., and J. G. Hovermale, 1968: An operational six-layer primitive equation model. J. Appl. Meteor., 7, 525-547.

Table 1. Potential predictors of maximum and minimum surface temperature for MOS screening regression at 60 and 72 hr. Numbers indicate valid time of predictors in hours after 0000 GMT. Stars indicate the predictor was smoothed by 5 points (*), 9 points (**), or 25 points (***). PE denotes the National Meteorological Center's primitive equation model (Shuman and Hovermale, 1968). TJ refers to the Trajectory model based on PE data (Reap, 1972). The 60-hr operational list gives those predictors which are currently being used.

Predictor	Tomorrow Night's Min (60 hr) (operational)	Tomorrow Night's Min (60 hr) (test)	Day after Tomorrow's Max (72 hr)
(a) PE Model			
850-mb height	48,48*	48,60	60,72*
500-mb height	36*,48,48*	48,60	60,72*
500-1000 mb thickness	48,48*	48,60,72*	60,72*,84*
850-1000 mb thickness	48,48*	48,60,72*	60,72*,84*
500-850 mb thickness	48,48*	48,60,72*	60,72*,84*
1000-mb temperature	48*,48**,48***	48**,60**,72**	60**,72**,84**
850-mb temperature	48*,48**,48***	48**,60**,72**	60**,72**,84**
700-mb temperature	–	48**,60**,72**	60**,72**,84**
Boundary layer potential temp	48*,48**,48***	48**,60**,72**	60**,72**,84**
Boundary layer U wind	48*,48**,48***	48*,60*,72*	60*,72*,84*
Boundary layer V wind	48*,48**,48***	48*,60*,72*	60*,72*,84*
Boundary layer wind speed	48*,48**	60*	72*
Boundary layer vertical velocity	–	60**	72**
850-mb U wind	24***	48*,60*,72*	60*,72*,84*
850-mb V wind	24***	48*,60*,72*	60*,72*,84*
700-mb U wind	24***	–	–
700-mb V wind	24***	–	–
1000-mb relative vorticity	48***	60**	72**
850-mb relative vorticity	48**	60**	72**
500-mb relative vorticity	48**,48***	60**	72**
850-mb vertical velocity	–	60**	72**
650-mb vertical velocity	–	60**	72**
Stability (700-1000 mb temp)	–	48**,60**,72**	60**,72**,84**
Stability (500-850 mb temp)	–	48**,60**,72**	60**,72**,84**
400-1000 mb mean rel hum	48**,48***	48**,60**,72**	60**,72**,84**
Precipitable Water	42**,42***	48**,60**,72**	60**,72**,84**
Boundary layer wind div	48**,48***	48**,60**,72**	60**,72**,84**
Boundary layer rel hum	–	48**,60**,72**	60**,72**,84**
Relative humidity (top of B.L.-720mb)	–	48**,60**,72**	60**,72**,84**
Relative humidity (720 mb-490 mb)	–	48**,60**,72**	60**,72**,84**
850-mb temperature advection	–	48**,60**,72**	60**,72**,84**
500-mb geostrophic vorticity advection	–	48**,60**,72**	60**,72**,84**
(b) TJ model			
Surface temperature	24**,24***	–	–
850-mb temperature	24**,24***	–	–
700-mb temperature	24**,24***	–	–
Surface dew point	24***	–	–
850-mb dew point	24***	–	–
700-mb dew point	24***	–	–
700-mb surface mean rel hum	24***	–	–
850-mb 12-hr net vert displ	24***	–	–
850-mb 24-hr net vert displ	24***	–	–
700-mb 12-hr net vert displ	24***	–	–
700-mb 24-hr net vert displ	24***	–	–
Surface 12-hr horiz conv	24***	–	–
850-mb 12-hr horiz conv	24***	–	–
George's K index	24***	–	–
(c) Other variables			
Sine day of year	00	00	00
Cosine day of year	00	00	00
Sine twice day of year	00	00	00
Cosine twice day of year	00	00	00

Table 2. Standard error of estimate (°F) and the reduction of variance on the dependent sample averaged at 49 test stations for the 60-hr and 72-hr forecast equations. These equations were developed from winter (December-February) data. Under the 60-hr equations, row A refers to the equations derived from the older predictor list and 6 years of developmental data. Row B represents equations developed from the identical predictor list as A but using 3 years of data. Row C indicates equations derived from 3 years of extended-range predictors. For the 72-hr equations, we used 3 years of data for the initial development and 4 years for the equations that are to be implemented.

Equation Set	Standard Error of Estimate (°F)	Reduction of Variance (%)
(a) 60-Hr		
Set A	6.8	62
Set B	6.4	64
Set C	6.3	65
(b) 72-Hr		
3-yr Developmental Sample	6.4	66
4-yr Developmental Sample	6.7	65

Table 3. The average at 49 test stations of the mean absolute error (°F) and the correlation coefficient between observed and forecast temperatures. The forecasts were made on independent data for the 1975-76 winter (December-February) season. The 60-hr forecasts are denoted the same as in Table 2. For the 72-hr forecasts, verifications on the same independent data (winter, 1975-76) are presented for both the MOS and the operational perfect prog system.

Equation Set	Mean Absolute Error (°F)	Correlation Coefficient
(a) 60-Hr		
Set A	5.8	.76
Set B	6.2	.73
Set C	6.1	.73
(b) 72-Hr		
Perfect Prog	6.4	.74
MOS (3-yr Developmental Sample)	6.5	.74

Table 4. Error Distribution of the MOS temperature forecasts made for 49 stations in the United States from December 1975 through February 1976.

Category 1: $T_{FCST} - T_{OBS} \leq -10°F$;

Category 2: $-10°F \leq T_{FCST} - T_{OBS} \leq 9°F$

Category 3: $T_{FCST} - T_{OBS} > 9°F$

Forecast A: operational 60-hr equations: 6 yr of developmental data
Forecast B: 60-hr equations, operational predictor list: 3 yr of developmental data
Forecast C: 60-hr equations, experimental predictor list: 3 yr of developmental data.

Equation Set	Number of forecasts in Category 1	2	3	Mean Algebraic Error (°F)
(a) 60-Hr				
A	327	3676	406	0.0
B	350	3564	495	0.3
C	360	3570	479	0.0
(b) 72-Hr				
Perfect Prog	458	3474	478	-0.3
MOS	565	3490	355	-1.5

Table 5. The ten most important predictors in the 72-hr maximum temperature equations for the winter (a) and spring (b) seasons. The predictors were ranked on the basis of a weighted scoring system in which the first predictor in an equation was assigned a score of 10 points, the second predictor was given 9 points, and so on until the final predictor in an equation received 1 point. Thus, an important predictor scored high because it was chosen frequently in the equations and because it was picked as a leading predictor in these equations. All model predictors were taken from the primitive equation model.

Order	Predictor
(a) Winter Equations	
1	850-mb temperature
2	Cosine 2 * day of year
3	Cosine day of year
4	Boundary layer potential temp
5	850-1000 mb thickness
6	850-mb V wind
7	Boundary layer U wind
8	1000-mb temperature
9	Boundary layer relative humidity
10	Layer 1 (top of boundary layer to 720 mb) relative humidity
(b) Spring Equations	
1	850-mb temperature
2	Cosine day of year
3	850-1000 mb thickness
4	Sine 2 * day of year
5	850-mb temperature advection
6	Boundary layer U wind
7	Layer 2 (720 mb to 490 mb) relative humidity
8	500-1000 mb thickness
9	500-mb geostrophic vorticity advection
10	Boundary layer potential temperature

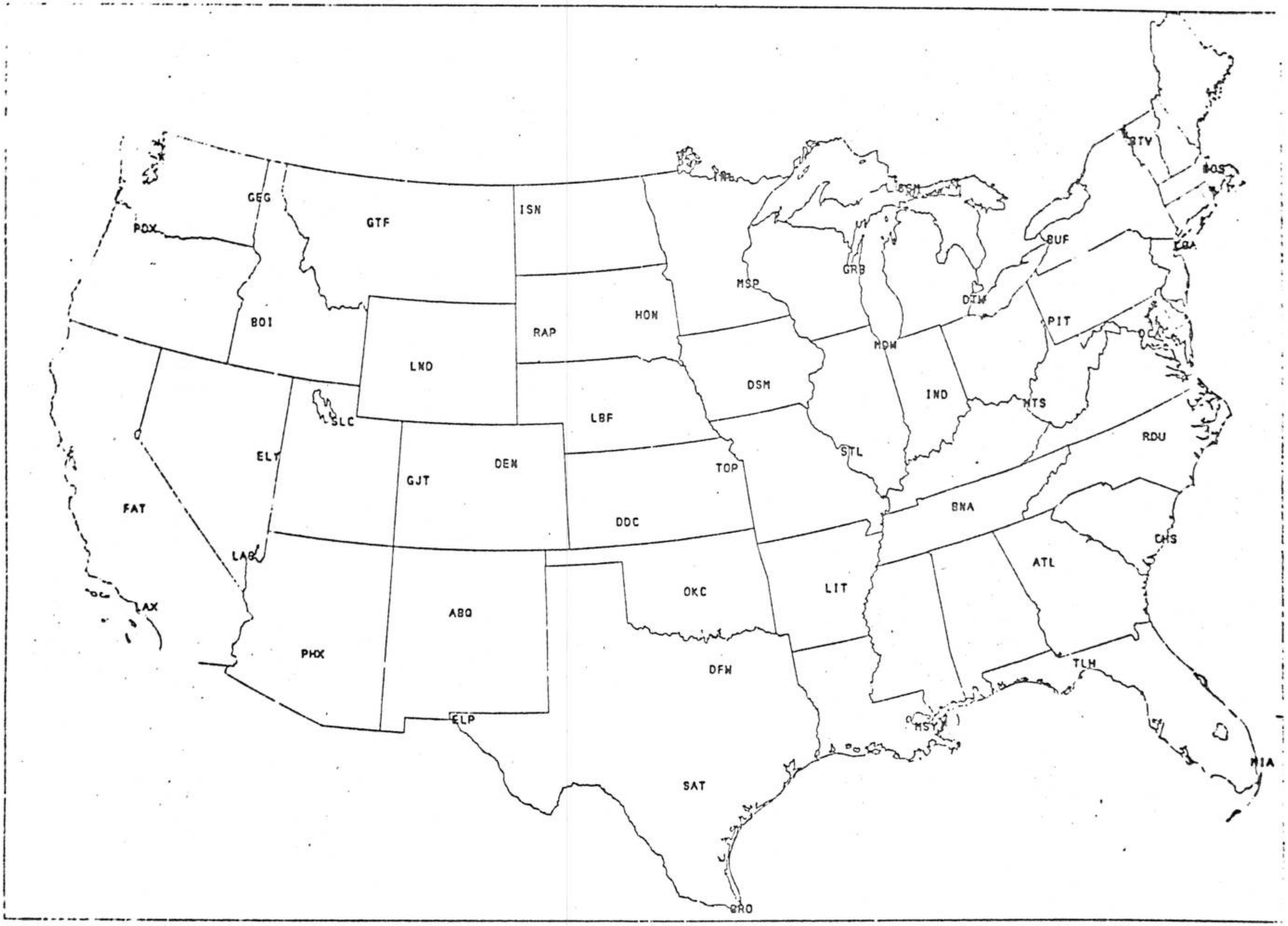

FIGURE 1. 49 STATION TEST NETWORK FOR 60 AND 72 HR FCSTS